Pizza course Reproducible and interactive data products and reports in R

Pizzacursus / OET sprintsessie

Willem Stolte, Peter Herman & Fedor Baart
Pizza course "Reproducible and interactive data products and reports in R"

When: 1st March 2016, 16:00 - 20:00 Where: Room: Ambitioin at Deltares, Delft

Many reports build heavily on complex analysis of data. Reproducibility is important for quality assurance (DQMS) and the use of reproducible reporting methods can (partly) automate the writing of such reports.

Reproducible reporting means here that the sources (scripts) for data analyzes, figures and tables are managed together with the source of the formatted text into one source document. The source document can be executed and published as pdf, word or (interactive) html.

Other recent initiatives supporting R as a tool for reproducible and interactive data analysis:

* ICES [Training course in the R environment] (http://ices.dk/news-and-events/Training/Pages/R-environment.aspx)
* R Shiny (including dashboard) is adopted by wikimedia to monitor the performance of their servers.

The course is meant for Data scientists (as instructors) with knowledge of data crunching, quality control and regridding and emerging datascientists (Participants) with an interest in R.

The course connects to existing projects which profit from interactive and / or reproducible data (re)processing, e.g.

* KPP Unlocking Marine Project Monitoring data
* Pilot digital water information system
* Sea Level Monitor
* Beheerbibliotheek kust
* Visualizations on existing data sources (Delta Data Portal, FEWS archive)

Intended content of the course:

1. Reproducible working methods with reproducible or dynamic documents
2. Interactive analysis of North Sea WQ data
3. Regridding of point data

Background material

* Learn to work with dynamic documents:
    * [R markdown reports and presentations] (http://rmarkdown.rstudio.com/)
    * [R notebooks] (http://rmarkdown.rstudio.com/r_notebook_format.html)
    * Jupyter and Conda for R (https://www.continuum.io/blog/developer/jupyter-and-conda-r?mc_cid=c91ccee4a5&mc_eid=3a45b30e21)
    * Rstudio webinars
* Interactive statistical models:
    * [Shiny integration in R markdown Documents] (http://rmarkdown.rstudio.com/authoring_shiny.html)
    * [Shiny apps] (http://shiny.rstudio.com/)
    * [Shiny dashboard] (https://rstudio.github.io/shinydashboard/)
* Global raster data reprocessing:
    * Examples from satellite measurements (swaths -> grids, ocean / north sea example)
    * Examples from radar measurements (regridden precipitation example)
    * Examples stations (regridden temperature raster)
    * Sample with a model (HIRLAM example)

Examples:

R markdown document regridding of bird counts

Interactive statistical models

Shiny apps at NIOZ